Initials & Surname: **Student Number**:

Date: 12 November 2015

EKN4815: Econometric Model Building
November Exam

Instruction: The exam is worth 100 marks. Working in this document by filling in answers to questions and reporting estimation results from Stata in the space provided. Time allowed is 3 hours.

QUESTION 1 (20 marks)

(a) Consider a model

$$Y_i = \beta_1 + \beta_2 X_i + u_i;\; i = 1, 2, \ldots, n$$

where Y_i is a binary variable that takes the value of 1 if the event takes place and 0 otherwise, and $\mathrm{E}(u_i) = 0$ for $i = 1, 2, \ldots, n$.

i. Explain fully the problems which arise if the above model is estimated by ordinary least squares (OLS).

(5 marks)

----

ii. How would you estimate the model . Discuss the advantages and disadvantages of this procedure.

(5 marks)

(b) A researcher wants to examine the newspaper reading habits of households. For this she collects data on fifty households and defines

$Y_i = 1$ if the i-th household purchases a newspaper, and $Y_i = 0$ otherwise.

She estimated the model defining $Y_i = f(S_i, E_i) + u_i$, where S_i is the years spent by the head of the i-th household in full time education, E_i is the average earnings of the head of the i-th household, and u_i is an unobserved disturbance term. The model was estimated by logit with the following results:

$$\begin{aligned} \hat{Y}_i \;\; &= \;\; \underset{(1.57)}{-2.56} + \underset{(0.10)}{0.521 S_i} + \underset{(0.012)}{0.067 E_i}; \;\; \log L_U = -321.25 \quad \log L_R = -416.01 \end{aligned}$$

where asymptotic standard errors are in brackets, $\log L_U$ is the log likelihood from the unrestricted model, and $\log L_R$ is the log likelihood of the model where all the slope coefficients are restricted to zero.

i. Explain how the coefficients were estimated.

(6 marks)

----

ii. Test the null hypothesis that all the slope coefficients are jointly equal to zero at the 5% level of significance.

(4 marks)

QUESTION 2 (20 marks)

(a) Consider a model

$$Y_{it} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit} + \alpha_i + u_{it};\ i = 1, 2, \ldots, n;\ t = 1, 2, \ldots, T$$

where Y is the dependent variable, X_j, $j = 2, \ldots, K$, are observed explanatory variables, α_i is an unobserved fixed effect and u_{it} is the disturbance term assumed to satisfy the usual regression model conditions. The index i refers to cross-section and the index t refers to the time period.

In the context of the model above briefly explain the least squares dummy variables method of estimation and its drawback.

(8 marks)

(b) The following estimates were obtained using an annually recorded US data panel of 550 individuals' wages over a period of 7 years. The dependent variable is the natural log of wage.

Independent variables	Pooled OLS	Random Effects	Fixed Effects
Years of full time education	0.091 (0.005)	0.092 (0.011)	
Black	−0.139 (0.024)	−0.139 (0.048)	
Hispanic	0.016 (0.021)	0.022 (0.043)	
Work experience	0.067 (0.014)	0.106 (0.015)	
Experience squared	−0.0024 (0.0008)	−0.0047 (0.0007)	−0.0052 (0.0007)
Married	0.108 (0.016)	0.064 (0.017)	0.047 (0.018)
Union membership	0.182 (0.017)	0.106 (0.018)	0.080 (0.019)

Six year dummy variables and a constant term were included in all three equations but the results are not reported. Black, Hispanic, Married and Union membership are dummy variables which take the value of one if the respondent has the relevant characteristic and are zero otherwise. The numbers in parentheses are standard errors of the coefficient estimates.

i. Explain why there are no fixed effects estimated coefficients for the first four explanatory variables in the table.

(3 marks)

----

ii. What interpretation would you give to the unobserved effects in a wage equation of this kind?

(3 marks)

iii. The OLS coefficients for Union membership and Married are higher than for the other two estimates. What does this suggest about the correlation between being married and being a member of a union and the unobserved effects? Explain your answer.

(3 marks)

iv. If there is a significant correlation between these two explanatory variables and the unobserved effects, what does this indicate about the properties of these estimates?

(3 marks)

QUESTION 3 (30 marks)

I- 10 marks

Consider the simple regression model

$$y = \beta_0 + \beta_1 x + u$$

and let z be a *binary* instrumental variable for x.

show that the IV estimator $\hat{\beta}_1$ can be written as $\hat{\beta}_1 = (\bar{y}_1 - \bar{y}_0)/(\bar{x}_1 - \bar{x}_0)$,

where $\bar{y}_0$ and $\bar{x}_0$ are the sample averages of y_i and x_i over the part of the sample with $z_i = 0$, and where $\bar{y}_1$ and $\bar{x}_1$ are the sample averages of y_i and x_i over the part of the sample with $z_i = 1$.

II- 20 marks

The data in FERTIL2.RAW includes, for women in Botswana during 1988, information on number of children, years of education, age, and religious and economic status variables.

(i) Estimate the model

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u$$

by OLS, and interpret the estimates. In particular, holding *age* fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many fewer children are they expected to have?

(5 Marks)

(ii) The variable *frsthalf* is a dummy variable equal to one if the woman was born during the first six months of the year. Assuming that *frsthalf* is uncorrelated with the error term from part (i), show that *frsthalf* is a reasonable IV candidate for *educ*. (*Hint:* You need to do a regression.)

(5 marks)

(iii) Estimate the model from part (i) by using *frsthalf* as an IV for *educ*. Compare the estimated effect of education with the OLS estimate from part (i).

(5 marks)

(iv) Add the binary variables *electric*, *tv*, and *bicycle* to the model and assume these are exogenous. Estimate the equation by OLS and 2SLS and compare the estimated coefficients on *educ*. Interpret the coefficient on *tv* and explain why television ownership has a negative effect on fertility.

(5 marks)

QUESTION 4 (30 marks)

Part A (5 marks)

Let mvp_i be the marginal value product for worker i, which is the price of a firm's good multiplied by the marginal product of the worker. Assume that

$$\log(mvp_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i$$
$$wage_i = \max(mvp_i, minwage_i),$$

where the explanatory variables include education, experience, and so on, and $minwage_i$ is the minimum wage relevant for person i. Write $\log(wage_i)$ in terms of $\log(mvp_i)$ and $\log(minwage_i)$.

Part B- (10 marks)

(Requires calculus) Let *patents* be the number of patents applied for by a firm during a given year. Assume that the conditional expectation of *patents* given *sales* and *RD* is

$$E(patents|sales,RD) = \exp[\beta_0 + \beta_1 \log(sales) + \beta_2 RD + \beta_3 RD^2],$$

where *sales* is annual firm sales and *RD* is total spending on research and development over the past 10 years.

(i) How would you estimate the β_j? Justify your answer by discussing the nature of *patents*.

(4 marks)

(ii) How do you interpret β_1?

(3 marks)

(iii) Find the partial effect of RD on $\mathrm{E}(patents|sales, RD)$.

(3 marks)

Part C- (15 marks)

We used the data in FERTILI.RAW to estimate a linear model for kids, the number of children ever born to a woman.

(i) Estimate a poisson regression model for kids, using the following variables: *educ, age, age2, black, east, northern, west farm, othrural, town, smcity, y74, y76, y78, 780, y82* and *y84.*

(a) Interpret the coefficient on y82. (2 marks)

(b) Report the output (Coefficient estimates and their p-values, likelihood value, R-squared and sigma) in the table provided below. (3 marks)

Regressors	Poisson QML Estimates (P-values in brackets)
Likelihood value	
R-squared	
Sigma	

(ii) What is the estimated percentage difference in fertility between a black woman and a nonblack woman, holding other factors fixed?

(2 marks)

(iii) Obtain $\hat{\sigma}$. Is there evidence of over- or underdispersion?

(3 marks)

(iv) Compute the fitted values from the Poisson regression and obtain the R-squared as the squared correlation between $kids_i$ and $\widehat{kids}_i$. Compare this with the R-squared for the linear regression model.

(5 marks)

Good Luck!!!